DOCUMENT RESUME

ED 313 318 SP 031 338

AUTHOR Shinghal, Rajjan; Le Xuan, Albert
TITLE Using Conceptual Analysis To Build Knowledge Bases.
PUB DATE 89
NOTE 15p.; For related document, see SP 031 339.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Abstract Reasoning; Cognitive Structures; *Concept Formation; *Concept Teaching; *Discriminant Analysis; Expert Systems; *Generalization; *Learning Processes
IDENTIFIERS *Conceptual Analysis

ABSTRACT

This paper describes the methods and techniques called Conceptual Analysis (CA), a rigorous procedure to generate (without involuntary omissions and repetitions) knowledge bases for the development of knowledge-based systems. An introduction is given of CA and how it can be used to produce knowledge bases. A discussion is presented on what is meant by learning and teaching a concept, a rule, and a principle. An explanation is provided of some of the pre-requisites and basic theories behind CA. The importance of positive examples and negative examples is emphasized and a specific instance of how a conceptual analysis can be performed to generate a knowledge base is demonstrated. It is pointed out the generalization and discrimination are the two most important concepts in the field of education and training. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED313318

USING CONCEPTUAL ANALYSIS TO BUILD KNOWLEDGE BASES

Rajjan Shinghal

Albert Le Xuan

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A. Le Xuan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

SP031338

BEST COPY AVAILABLE

USING CONCEPTUAL ANALYSIS TO BUILD KNOWLEDGE BASES

Rajjan Shinghal
Concordia University, Computer Science Department, 1455 De Maisonneuve West
Montreal, Que H3G 1M8, Canada

And

Albert Le Xuan
Alphamega Kknowledge Engineering, 4987 Walkley Avenue, Montreal, Quebec H4V 2M3

1.0 ABSTRACT

It is our belief that generating the knowledge base is a fundamental step in the development of expert systems and learning systems.

An ever increasing number of educational technologists and knowledge enginers have become aware of the challenge of knowledge engineering and have searched for more ways of improving methods and techniques that will ensure the most efficient and effective acquisition of capabilities and knowledge on the part of the learner or user.

Quite often, however, despite the perfectly written objectives and the undoubted recognition that we know where we want to go, we feel unsure of having found methods and techniques to generate the knowledge base to build the "bridge" which will take our learners from where they are to where we want them to go. When the knowledge base is not "good", that is, when it is not accurate, up-to-date, complete, or easy to understand, then it does not serve its purpose.

This paper describes the methods and techniques called Conceptual Analysis, a rigorous procedure to generate - without involuntary omissions and repetitions - knowledge bases for the development of knowledge-based systems.

2.0 INTRODUCTION

In this paper we introduce the notion of Conceptual Analysis (CA) and how we can use it to produce knowledge bases. We discuss what is meant by learning and teaching a concept, rule, and principle. We explain some of the pre-requisites and basic theories behind CA. We emphasize the importance of positive examples and negative examples in learning and teaching concept and then show a specific instance of how a conceptual analysis can be performed to generate a knowledge base.

3.0 SOME THEORIES BEHIND CA

3.1 Generalization & Discrimination: Generalization and discrimination are the two most important concepts in the field of education and training. In fact, the aim of education is to create situations which will allow the student to generalize within a class and to discriminate between that class and other classes. In other words, to design materials which will allow the learner to generalize and to discriminate is the goal of instructional design.

Definitions: (i) To generalize within a class is to make the same response to different stimuli. To behave in the same manner in front of different situations is also gen-

eralize within a class of situations. To go beyond what is given is also to generalize within a class of objects, qualities, quantities, events and relations.

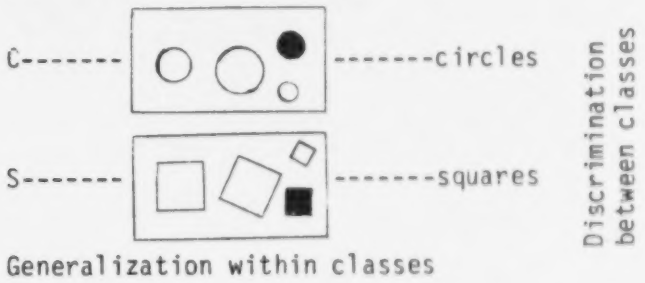
(ii)To discriminate between classes is to make different responses to different stimuli.

Negative & Positive Examples: (i) To say "doggy" in the presence of poodles, greyhounds, and pointers is to generalize within a class of things called "dogs".(ii) To say "bird" at the sight of ducks, pelicans, eagles is to generalize within a class of things. (iii) To say "dog" in the presence of horses is to generalize within a class of animals, but the generalization here is not a correct one.(iv) To run away from cobras but not from water snakes is to discriminate between two classes of objects.

3.2 When Can We Say That An Individual Has Learned A Concept? In other words, what must an individual do to show that he (for brevity, we shall use masculine pronouns for both genders) has learned a concept?

(i) He must be able to generalize within a class of objects, events or relations, and (ii) discriminate between that class and other classes.

Example 1: Consider the following classes of figures:



When an individual makes the same response "circle" to each of the figures of class (C), and the same response "square" to each member of class (S), he is generalizing within a class of circles and a class of squares; he is also discriminating between a class of circles and a class of squares. The psychologist would say that the individual has learned the concept of circle and the concept of square.

Example 2: When a baby is capable of distinguishing his dad from his uncle (who is his father's twin brother), we say that the baby has learned the concept of "dad", though he cannot say it in words.

Non-example 1: Paul said "square" in the presence of square figures and also "square" in the presence of figures which resemble squares but are not squares. Paul has learned to generalize but his generalization was incorrect. We say that Paul has not learned the concept of square.

Non-example 2: A dog has been trained to detect the enemy. After three months' training the dog was still unable to discriminate between friends and foes. We say that the dog has not learned the concept of enemy.

Non-example 3: When a person is capable of giving the definition of a term but is not able to give or recognize an example of the concept represented by the term, he has not learned the concept represented by the term.

3.3 How To Teach A Concept: We have analyzed the behavior of the learner to see what he does when he learns a concept: he generalizes within a class of stimuli, and he discriminates between that class and other classes. Now we are going to analyze the behavior of the instructor to see what he does when he teaches a concept, that is, how he goes about teaching the learner to generalize and to discriminate among things.

Procedure:

Suppose we have to teach the concepts of proposition and predicate, here is how we could proceed:

1) Tell the learner why he should learn these two concepts:

"Introduction: 'Proposition' and 'predicate' are the first two basic concepts of logic. In other words, in order to be able to understand and learn logic we must first learn these two concepts."

2) Show the learner one example proposition:

"The following sentence can be either true or false: 'The box is on the table'. It is a proposition."

3) As one cannot generalize with one single example, give the learner another example:

"The robot is by the window." This sentence can be either true or false but not both. It is a proposition."

4) To allow the learner to discriminate between 'proposition' and 'non-proposition', give him several non-examples of proposition.

"The following sentences can be neither true or false: Slow down! Please come early! They are not propositions."

5) Give him a definition of proposition:

"In logic a proposition is a sentence that can be either true or false."

6) Give him an example predicate:

"Consider the following proposition: 'The box in under the table'. It deals with two items of the real world: the box & table. The relationship between the box and the table is specified by the words is on. This part of the proposition which makes an assertion about the individuals (box & table) is called a predicate."

7) Give him several other examples of predicate:

"Other examples of predicate are: is by, is near."

8) If necessary, give him some examples of notation of individuals:

"For convenience in writing expressions, individuals are given one-word names:

BOX for the box
TABLE for the table, etc.

9) We can also give him some examples of notation of predicate:

"For convenience in writing expressions, we also give each predicate a one-word name:

ON	for	is on
NEAR	for	is near

Finally, the original proposition can be written in terms of the predicate and the individual:

ON(BOX,TABLE)

which is read 'the box is on the table'."

10) Give him a definition of predicate:

"Definition: That part of the proposition which makes an assertion about the individuals is called the predicate, and the items about which a proposition which makes an assertion are called individuals.

11) Perhaps it is time to give him a test to see if he has learned to generalize and discriminate within and among classes of propositions, predicates and individuals:

"Exercises: (a) BY(TABLE,WINDOW) is read _____

(b) Write in symbols 'the box is near the window' _____

(c) What is a proposition?

(d) What is a predicate, an individual?

Answers: (a) BY(TABLE,WINDOW) is read 'the table is near the window'.

(b) 'The box is near the window' is written in symbols:

NEAR(BOX,WINDOW).

(c) A proposition is a sentence that can be either true or false.

(d) That part of the proposition which makes an assertion about the individuals is called the predicate, and the items about which a proposition makes an assertion are called individuals."

3.4 Positive & Negative Examples: We have seen that:

(i) To teach a concept we must teach the learner to generalize within a class and to discriminate between that class and other classes, and

(ii) To teach the learner to generalize and to discriminate, we must present him at least two positive examples and two negative examples of the concept. The number of positive and negative examples necessary for these two operations depends on the prerequisites of the target audience and the complexity of the subject matter.

Now, where do these positive and negative examples come from? Usually, it is the subject matter expert who has to supply them.

Definitions: (i) The positive examples (or simply examples) of a concept or class are the members or elements of that class or concept; and (ii) the negative examples (or non-examples) of a class or concept are the members or elements of a class which is similar to but different from the class or concept under study.

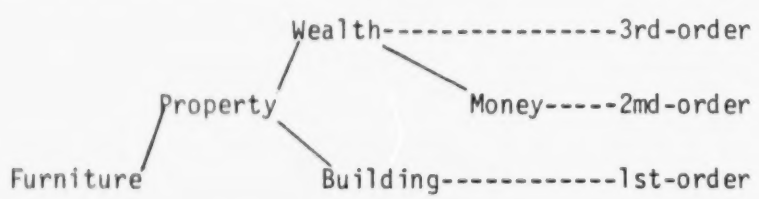
Positive & Negative Examples of Positive & Negative Examples: (i) A foxhound belongs to a class of objects called 'dog'. A foxhound is said to be a positive example (or example, examplar, or instance) of the concept (class) 'dog'. (ii) A fox which looks somewhat like a dog but is not a dog, would be a good negative example (or non-example) of the class 'dog'. (iii) An elephant which has a few common characteristics with a setter or pointer, would not be a good negative example of the concept 'dog'. (iv) Chair is an excellent negative example of 'stool', because chair and stool both possess the same attributes (properties) except one: a chair has a back whereas a stool does not. (v) The following are some negative examples - in descending order of subtlety- of the concept 'dog': (a) Jackals or foxes are negative examples of high subtlety, (b) cats are negative examples of average subtlety, (c) black and white are negative examples of obvious irrelevancy.

4.0 SOME PRE-REQUISITES TO CA

4.1 Primary, Secondary & Nth-order Concepts & Related Concepts: The ideas of primary, secondary, nth-order, coordinate, subordinate, supraordinate concepts are very important in knowledge engineering.

Definitions: (i) According to Skemp (1971), concepts that are derived from our sensory and motor experiences of the outside world are called primary concepts, and concepts that are derived from primary concepts are called secondary concepts. (ii) For the sake of clarity we will use the term "1st-order concept" for "primary concept", and "2nd-order,3rd-order, ... ,nth-order concepts," as the case may require, for "secondary concepts." (iii) Supra-ordinate or superordinate concepts are concepts that are of the highest order or level in a conceptual hierarchy or structure. (iv) Coordinate concepts are those of the same rank or order in a conceptual hierarchy. (v) Subordinate concepts are those that are of lower-order in a conceptual hierarchy.

Example: Below is a diagram of a conceptual hierarchy:



In this example, furniture and building are first-order concepts; they are also co-ordinate concepts. Property and money are 2nd-order concepts; they are also co-ordonate concepts. Wealth is the supra-ordinate or superordinate concept.

Note that furniture would become a "2nd-order concept" if it is broken down into chair, stool, bed. Furniture would become a "3rd-order concept" if chair is broken down into 'iron chair' and 'wooden chair.' But in general, 2nd-order concepts are those that are derived from 1st-order concepts; 3rd-order concepts are those that are derived from 4th-order concepts; 5th-order concepts are those that are derived from 4th-order concepts, and so on and so forth. Everything is relative.

4.2 Algorithms & Heuristics: In the study of problems, plans and operators are divided into two types: algorithms & heuristics. In non-AI programming, the programmer first develops a step-by-step method, called algorithm, for solving certain classes of problems. There are unfortunately, two difficulties with this approach:

(i) There are some types of problems for which there is no algorithm that will solve them, and

(ii) When there is an algorithm that will solve a certain type of problems, the algorithm may be practical only for the smallest excamples of the problems, and are so inefficient as to be out of the question for practically sized problems.

Definitions: (i) An algorithm is an exhaustice and resultative set of rules which, if followed, will automatically generate the correct solution. (ii) A heuristic is any rule of thumb or probabilistic procedure that helps us discover a solution to a problem.

Positive and Negative Examples: (1) The rules for multiplication constitute an algorithm: if we use them properly, we always get the right answer. (ii) Chess manuals do not give a prescription guaranteed to lead to success. Rather, they contain such rules as:

(a) A bird in the hand is worth two in the bush.

(b) A stitch in time saves nine.

(c) Look before you leap.

Note that each proverb is a good advice. That is, each is certainly worth considering in the appropriate situation. Unfortunately, no proverb is guaranteed to always yield desirable results. Nevertheless, heuristic search is the principal problem-solving technique of Artificial Intelligence.

Measurement of Procedure Acquisition: To show that one has acquired the concepts of algorithm and heuristic, one must be able to:

(i) Discriminate between the class of algorithms & the class of heuristics.

(ii) Determine if the task is well-defined: if the terms are vague, one cannot apply the algorithmic procedure.

(iii) Determine if the task is manageable, that is, one must be aware of combinatorial explosions. For instance, the 10^{120} possible moves of a chess game: the fastest computer today could not explore them all in the time that remains before the death of the Sun.

(iv) Know when to use algorithmic procedures and when to use heuristics.

4.3 Knowledge and Skills in a Given Domain: What are knowledge and skills in a given domain or field? Dictionaries give such definitions as: "Knowledge in a given domain is what we know in this domain," and "Skill is the ability gained by practice," which is quite vague.

Definitions and Explanations: Knowledge in a given domain consists of descriptions, relationships, and procedures in that domains. In short, knowledge in a given domain consists of (i) the symbolic descriptions and relationships in that domain and (ii) the symbolic descriptions of procedures for manipulating these descriptions. The descriptions in a knowledge base are sentences in some language who elementary

components consists of concepts, rules, principles, facts, and procedures for applying and interpreting descriptions in specific applications.

Examples of Knowledge: The following are some examples of knowledge:

(i) Symbolic descriptions of concepts, rules or procedures used to infer situations from observables.

(ii) Symbolic descriptions of rules and procedures used to infer likely consequences from given situations.

(iii) Symbolic descriptions of rules and procedures used to infer system malfunctions from observables.

(iv) Symbolic descriptions of concepts, rules, procedures used to prescribe remedies for malfunctions.

Examples of Skills: The following are some examples of skills:

(i) A mechanic who can repair an engine in two hours instead of ten is said to have acquired great skill.

(ii) A person who is capable of applying knowledge to produce solutions, both efficiently and effectively, is said to be very skillful.

5.0 USING CA TO GENERATE KNOWLEDGE BASES

The knowledge base is the raw material with which we build knowledge-based systems. It is made up of at least two components: the declarative knowledge base which contains concepts, rules and principles, facts and opinions, and the procedural knowledge base which contains procedures to perform such tasks ad doing an experiment to illustrate scientific concepts, starting a car in the morning, or changing a tyre.

The following is the CA the authors have performed to generate the knowledge base for the development of the structure (lesson), RELATIONS, one of the component parts of the course, SOME MATHEMATICAL PRE-REQUSITES TO CYBERNETICS. The knowledge bases for the rest of the course were generated in the same manner. This CA has been performed from a term (RELATIONS) but a CA can also be performed from any stimulus: a phrase. sentence, or text.

5.1 Procedure

CA : RELATIONS
Part A : Input-Output Inventory

Input	Outputs
1.0 Relations →	2.0 Introduction: The concept of relation is very important both in everyday life and in science. Without the concept of relation we will not be able to describe or explain anything.
↘	3.0 Definition: A relation is something that connects objects. It can be likened to cement or glue which holds things together, but it is not something we can see, touch or photograph. It can be represented by a verb or verbal phrase. Other names for 'relation' are: link, correspondence, dependence, function, associatation, and so forth.

1.0 Relations →

4.0 Positive Examples

4.1 Consider the sentence: "Paul loves Jane." The verb "loves" which related Paul to Jane is a relation.

4.2 The following are some other examples of relations:

... is north of ...
... depends on ...
... borrows ...
... borrows ... from ...
... is equal to ...
... is between ... and ...

5.0 Negative Examples

5.1 Consider the following group of words: "Paul beautiful Jane." We cannot say anything about Paul and beautiful Jane because of the absence of such verbs and verbal phrases as:

... is in love with ...
... likes ...
... hates ...,etc.

which express a certain link between objects.

5.2 Consider the sentence: "John borrows money from the bank." Now, if we strike out the verb "borrows", this group of words would be meaningless. In other words, without the concept of relation we cannot describe or explain anything.

6.0 Arrow Diagrams

6.1 Consider the relation "... is the brother of ..." in a set of people: "John is the brother of Helen." We can represent (visualize) this relation by drawing an arrow starting from John to Helen like this:

John → Helen or like this: Helen ← John.

The important here is that the arrow goes from John to Helen No arrows go from Helen to John because Helen is not the brother of John.

6.2 Consider the relation "... is the brother of ..." in a set of two males: "John is the brother of Paul." To represent this relation in a set of two males, we draw double arrows like this:

John ⇄ Paul or like this: John ⟷ Paul.

1.0 Relations ⟶ 7.0 Some Related Rules

7.1 One cannot be conscious of one thing only. In order to have one thing, we must have at least three: a thing, a relation, and another thing. To explain what "uncle" is, for instance, we must have three things:

1. Uncle (the thing to be explained)
2. is the brother of (a relation)
3. One's father or mother (another thing)

7.2 Two unrelated terms cannot exclude each other (Hamelin,1907). In other words, the excluded thing must also be present before we can have an exclusion. To suppress one term, the other will disappear just as shadow will disappear when light vanishes.

8.0 Exercises

(1) Underline the relations in the following:

a. Mark is as tall as John.
b. 4 is the square root of 16.
c. 16 is the square of 4.
d. Ottawa is west of Montreal.

(2) Which of the following is/are true?

a. We cannot teach a subject without the concept of relation.
b. Relations are something that holds things together.
c. A relation can be expressed by a verb.
d. A relation can be visualized by an arrow.

9.0 Answers

(1) You should have underlined: is as tall as, is the square root of, is the square of, is west of.

(2) All are true.

2.0 Introduction: ⟶ NTBA Short for Not To Be Analyzed. This means that by convention we will not analyze (breakdown) the content of this item.

The concept ...

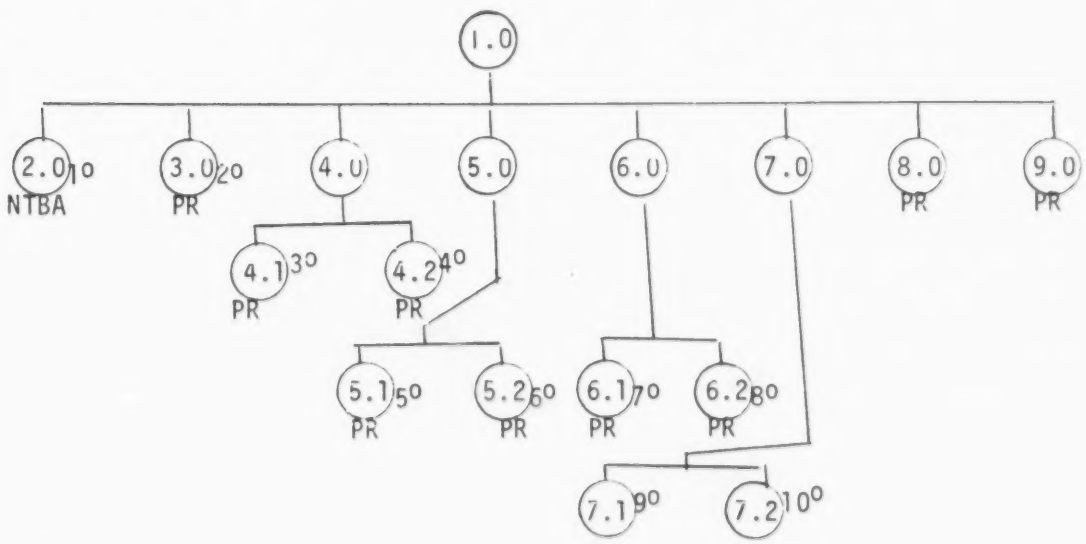
3.0 Definition ⟶ PR Short for Pre-Requisite, by which we mean that this item contains no concepts which call for further breakdown.

A relation ...

CA : RELATIONS
Part A : Input-Output Inventory

4.0 Positive Examples	
4.1 Consider ... ⟶	PR
4.2 The ... ⟶	PR
5.0 Negative Examples	
5.1 Consider ... ⟶	PR
5.2 Consider ... ⟶	PR
6.0 Arrow Diagrams	
6.1 Consider ... ⟶	PR
6.2 Consider ... ⟶	PR
7.0 Some Related Rules	
7.1 One ... ⟶	PR
7.2 Two ... ⟶	PR
8.0 Exercises ⟶	PR
9.0 Answers ⟶	PR

CA : RELATIONS
Part B : Representation of CA



CA : RELATIONS
Part C : Sequential Index

1° (2.0 - 1.0) Introduction: The concept ... anything.

2° (3.0 - 1.0) Definition: A relation ...

3° (4.1 - 4.0 - 1.0) Positive Example: Consider ...

4° (4.2 - 4.0 - 1.0) Positive Example: Consider ...

5° (5.1 - 5.0 - 1.0) Negative Example: Consider ...

6° (5.2 - 5.0 - 1.0) Negative Example: Consider ...

7° (6.1 - 6.0 - 1.0) Arrow Diagram : Consider ...

8° (6.2 - 6.0 - 1.0) Arrow Diagram : Consider ...

9° (7.1 - 7.0 - 1.0) Related Rule : One ...

10° (7.2 - 7.0 - 1.0) Related Rule : Two ...

11° (8.0 - 1.0) Exercises : (1) Underline ...

12° (9.0 - 1.0) Answers : (1) You ...

5.2 How Was Part A Generated: Each input is written on the left side of the page. It generates one or more outputs which appear(s) on the right side of the page. Each of these outputs will in turn become new inputs for further outputs and so on, until we reach outputs which call for no further analysis: these are the pre-requisites.

In general, each new concept generates a structure which is composed of such element as the following:

1) An introduction
2) A definition
3) Some positive examples (or some experiments if it is a science concept)
4) Some negative examples
5) A procedure (if necessary)
6) Some related concepts
7) Some related rules
8) Exercises with answers
9) References & bibliography,etc

Rule: Each output (except those that are followed by the symbol PR) is transferred to the left to become a new input which will generate one or more outputs.

To generate Part A of the analysis we proceeded as follows:

1) We divided all the pages into two parts, forming two columns: one for inputs and one for outputs.

2) We recorded the material to be analyzed (in this case, the term relation) in the input column and numbered it 1.0.

3) Input 1.0 produced eight outputs: 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, and 9.0.

4) Next, we drew a horizontal line under the first input & its eight outputs.

5) Then, we brought down output 2.0 (introduction) in the input column to turn it into an input, and recorded it under the line we had just drawn.

6) Input 2.0 produced NTBA, short for Not To Be Analyzed. This means that, by convention, we will not analyze (breakdown) this item as it contains concepts which are analyzed elsewhere.

7) We drew another horizontal line.

8) We brought down output 3.0 in the input column to turn it into an input and wrote it under the line we had just drawn.

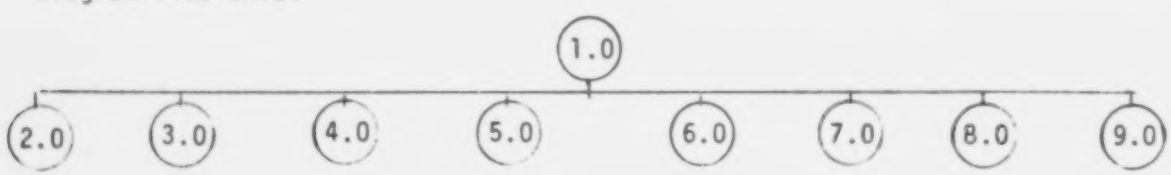
9) Input 3.0 produced PR, short for Pre-Requisite, by which we mean that we will not pursue our analysis any further, assuming that all the concepts in this item are known to the target population.

10) We drew another horizontal line.

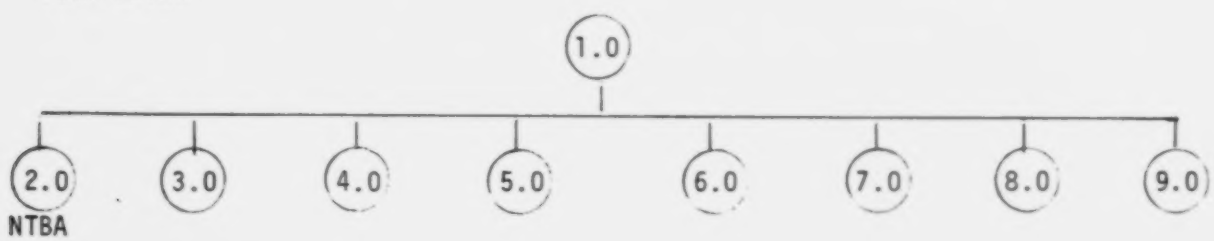
11) We brought down output 4.0 in the input column to turn it into an input and treated in the same manner until all the outputs were turned into inputs and produced PRs.

5.3 How Was Part B Generated: To reorganize the input-output inventory into a hierarchy or structure we proceeded as follows:

1) As one can see, input 1.0 produced eight outputs. To show this, we drew a tree diagram like this:



2) As input 2.0 produced NTBA, we added to the partital tree diagram which now looks like this:



3) We proceeded in the same manner until we exhausted all the inputs and outputs of the inventory and obtained the final tree diagram.

4) The symbols 1^{o}, 2^{o}, 3^{o}, etc represent the order of the component parts of the hierarchy. The ordering begins with the branches, from left to right, and from the bottom upwards.

5) As one can see, the CA begins at the top and moves downwards. But the concepts, examples or rules, represented by the nodes, are ordered from the bottom upwards. This means that the analysis proceeds from the higher-order concepts to the lower-order concepts, and the presentation of the concepts, examples and rules, proceeds from the lower-order concepts to the higher-order concepts.

6) Note: When the domain is complex the tree diagram may accupy quite a large space. In that case it is necessary to break the tree diagram into as many branches as possible. Some tree diagrams can give a sense of awe, but without the help of tree diagrams we will have to resort to intuition and guesswork which are usually more time-consuming.

5.4 How Was Part C Generated: To generate Part C we proceeded as follows:

1) We picked out the first item (1^{o}) on the tree diagram, then looked at the input-output inventory, and copied the content of 1^{o} on a separate sheet of paper. This is the first element of the hierarchy.

2) Next, we picked out the second item (2^{o}) and copied its content. We proceeded in the same manner until the last item was disposed of. The end-result is the knowledge base with which we built the course.

6.0 REFERENCES & BIBLIOGRAPHY

1) Feigenbaum, E.A. & McCorduck, Pamela (1983). Fifth Generation. Reading, MA: Addison-Wesley Publishing & Company.

2) Hamelin (1907). Les Eléments Principaux de la Représentation. Paris: Alcan.

3) Le Xuan & Chassain, J.C. (1975). Analyse Comportementale. Paris: Fernand Nathan.

4) _____ (1985). Topdown Conceptual Analysis. Montreal: Pedagogical & Technological Innovations.

5) Skemp, R.R. (1971). The Psychology of Learning Mathematics. London: Penguin Booksé

6) Walters, J.R. & Nielson, N.R. (1988). Crafting Knowledge-based Systems. New York: John Wiley & Sons.

7) Waterman D.A. (1986). A Guide to Expert Systems. Reading, MA: Addison-Wesley Publishing Company.